REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
20-10-2013	Final Report	1-Aug-2010 - 31-Jul-2013

4. TITLE AND SUBTITLE Final Report on ARO Research Grant W911NF-10-1-0037 “Structure Inference from Mobility Encounters”	5a. CONTRACT NUMBER W911NF-10-1-0337
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS Leonidas J. Guibas	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Stanford University Research Management Group 3172 Porter Drive Palo Alto, CA 94304 -1212	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 58000-MA.17

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

In the final year of the great we have focused our efforts on the large-scale analysis of mobility data, such as GPS traces from vehicles.

Understanding trajectory data sets, and extracting meaningful information from them, entails many computational challenges, from data set size to sensing uncertainty and trajectory heterogeneity in quality, format, and temporal

15. SUBJECT TERMS

mobility, GPS traces, optimization, Makrov decision process, geometric data analysis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Leonidas Guibas
UU	UU	UU	UU		19b. TELEPHONE NUMBER 650-723-0304

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report on ARO Research Grant W911NF-10-1-0037 “Structure Inference from Mobility Encounters”

ABSTRACT

In the final year of the great we have focused our efforts on the large-scale analysis of mobility data, such as GPS traces from vehicles.

Understanding trajectory data sets, and extracting meaningful information from them, entails many computational challenges, from data set size to sensing uncertainty and trajectory heterogeneity in quality, format, and temporal support. At the same time, individual trajectories can have complex shapes, and even small nuances can make big differences in their semantics. A major tension in understanding trajectory data is between the need to capture the fine details and shape features of individual trajectories and the ability to exploit the wisdom of the collection, i.e., to take advantage of the information embedded in a large collection of trajectories but missing in any individual trajectory. This emphasis on the wisdom of the collection is one of the main novelties of the work presented.

We discuss results on extracting a pathlet dictionary for a trajectory collection, on exploiting a collection to better map individual trajectories to an underlying road network, and on exploiting such a collection to derive information that helps the mobile entities.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
10/20/2013	16.00	Mo Li, Yuan Yao, Xiaoye Jiang, Leonidas Guibas. Property management in wireless sensor networks with overcomplete radon bases, ACM Transactions on Sensor Networks, (05 2013): 0. doi: 10.1145/2480730.2480739
10/29/2012	9.00	Lek-Heng Lim, Yuan Yao, Yinyu Ye, Xiaoye Jiang. Statistical ranking and combinatorial Hodge theory, Mathematical Programming, (11 2010): 0. doi: 10.1007/s10107-010-0419-x

TOTAL: 2

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
03/11/2012	3.00	Frederic Chazal, Daniel Chen, Leonidas Guibas, Xiaoye Jiang, Christian Sommer. Data-Driven Trajectory Smoothing, 19th ACM\ SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2011/11/01 03:00:00, . : ,
10/29/2012	8.00	D. Chen, L. Guibas, J. Hershberger, J. Sun. Road Network Reconstruction for Organizing Paths, ACM-SIAM Symposium on Discrete Algorithms (SODA). 2010/01/25 03:00:00, . : ,
TOTAL:		**2**

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
03/11/2012	2.00	Xiaoye Jiang, Mo Li, Yuan Yao, Leonidas J. Guibas. Overcomplete Radon Bases for Target Property Management in Sensor Networks, 10th International Conference on Information Processing in Sensor Networks (IPSN). 2011/04/12 03:00:00, . : ,
03/11/2012	5.00	Xiaoye Jiang, Yuan Yao, Han Liu, Leonidas Guibas. Detecting Network Cliques with Radon Basis Pursuit, International Conference on Artificial Intelligence and Statistics (AISTATS). 2012/04/21 03:00:00, . : ,
03/11/2012	4.00	Xiaoye Jiang, Jonathan Huang, Leonidas Guibas. Fourier-Information Duality in the Identity Management Problem, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. 2011/09/05 03:00:00, . : ,
09/10/2013	10.00	Mridul Aanjaneya, Frederic Chazal, Daniel Chen, Marc Glisse, Leonidas Guibas, Dmitriy Morozov. Metric Graph Reconstruction from Noisy Data, 27th Annual ACM Sympoium on Computational Geometry, 2011. 2011/06/13 03:00:00, . : ,
09/10/2013	11.00	Daniel Chen, Anne Driemel, Leonidas Guibas, Andy Nguyen. Approximate Map Matching with respect to the Fréchet Distance, ALENEX 2011. 2011/01/24 03:00:00, . : ,
09/10/2013	12.00	Mo Li, Xiaoye Jiang, Leonidas Guibas. Fingerprinting Mobile User Positions in Sensor Networks, Thirtieth International Conference on Distributed Computing Systems (IEEE ICDCS). 2010/07/21 03:00:00, . : ,
10/20/2013	13.00	Yang Li, Qixing Huang, Michael Kerber, Lin Zhang, Leonidas Guibas. Large-Scale Joint Map Matching of GPS Traces, Proc. 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2013/11/07 03:00:00, . : ,
10/20/2013	14.00	Hao Su, Chen Chen, Qixing Huang, Lin Zhang, Leonidas Guibas. Pathlet Learning for Compressing and Planning Trajectories, Proceedings of the 21th SIGSPATIAL International Conference on Advances in Geographic Information Systems,. 2013/11/07 03:00:00, . : ,
10/20/2013	15.00	Haochen Tang, Michael Kerber, Qixing Huang, Leonidas Guibas. Locating Lucrative Passengers for Taxicab Drivers, Proceedings of the 21th SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2013/11/07 03:00:00, . : ,

TOTAL: **9**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

Best paper award at the Eurographics Symposium on Geometry Processing, 2013.

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Chen Chen	0.25	
FTE Equivalent:	**0.25**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Qixing Huang	0.07
Sharath Sharathkumar	0.20
FTE Equivalent:	**0.27**
Total Number:	**2**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Leonidas Guibas	0.07	
FTE Equivalent:	**0.07**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	Discipline
Haque Muhammad Ishfaq	0.10	Computer and Computational Sciences
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 1.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 1.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 1.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 1.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME	
Total Number:	

Names of personnel receiving PHDs

NAME	
Daniel Chen	
Jonathan Jiang	
Total Number:	**2**

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

In this last year of the grant, we have focused our efforts on the analysis of large collections of GPS traces, with goal of extracting shared structure in the collection and exploiting that to improve individual trajectories (e.g., in matching them to a map), or in understanding various more semantic aspects of the environment, based on the trajectories. We were fortunate to have access to a very large set of taxicab traces from Beijing for this work.

I. Pathlet Learning for Trajectory Compression and Planning

The pervasiveness of GPS devices has created many large datasets of pedestrian and vehicle trajectories. Compressing such large data sets is obviously of interest. Furthermore, compression is tightly coupled with shared structure extraction, related to the semantics of trajectories. Such higher-level trajectory understanding can benefit a variety of applications ranging from the study of population migration routes, vehicular traffic patterns, and the state of city road networks. In this work, we have sought to extract latent shared structure among many human trajectories.

Our work is motivated by the seminal work of Gonzalez et al, who discovered that human trajectories show a high degree of spatial and temporal regularity, i.e., human beings have high probability of repeating similar mobility patterns. We seek to find a set of path segments that have semantic meanings, referred to as the pathlet dictionary, out of which most trajectory can be reconstructed by concatenating a few of these segments. We note that this can also be viewed as a joint trajectory segmentation problem, where the pathlet structure becomes only apparent in the context of a trajectory collection.

We formulate pathlet learning as solving an integer linear program (ILP), whose objective function minimizes the size of the pathlet dictionary as well as the number of pathlets that are used to reconstruct each trajectory. To solve this ILP on large datasets, we introduce a decoupled approach, which optimizes a lower bound of the original objective function. We have tested our algorithm on a large-scale real world dataset, which contains 230K trajectories of taxi cabs in Beijing. Our algorithm extracts a pathlet dictionary containing around 130K pathlets, which can reconstruct all trajectories in the dataset using 7 pathlets on average per trajectory. This number is significantly smaller than the average number of edges used to represent trajectories on a road map (which is 60.1), or the average number of “turns” that might be provided in a navigation system to realize the trajectory (which is 36.7). More interesting than these global compression statistics is the semantic information one gleans about Beijing traffic through the pathlets. For example, in the downtown area, pathlets are shorter, corresponding to smaller commute ranges, while for the region near the airport (located in a suburban area), the pathlets are longer, corresponding to long trips from the city to the airport.

Our pathlet dictionary we described can be useful not only for trajectory compression but in trajectory synthesis applications as well. In that data set, frequently used pathlets in the dictionary represent driving segments chosen by many taxi cab drivers in Beijing, reflecting the joint wisdom of a highly skilled set of professionals. To show the usefulness of the extracted pathlets, we implemented a route planning application for the city of Beijing. Experimental results are competitive against those obtained from Google Maps and at times superior.

II. Large-Scale Joint Map Matching of GPS Traces

Map matching is the procedure of determining the path of a user on a map from a sequence of GPS positions of that user -- a trajectory. This procedure finds use in many mobility related application, such as urban traffic modeling, dynamic road map generation, and mobility pattern mining. Since collecting highly accurate GPS traces on a city scale is quite costly, most of the trajectory data available today were obtained indirectly through GPS-equipped vehicles or users with GPS-enabled cellular phones. The majority of the collected trajectories inevitably contain a large amount of uncertain and incomplete information. For example, one form of uncertainty comes from GPS noise, which is particularly severe in urban environments due to signals obstructed by or reflected off buildings (urban canyons). Incomplete data is often the result of a low sampling rate, due to limitations on storage and communication bandwidth. For instance, 50% of the Beijing Taxi Trajectories we employed in this study have at most one sample per minute.

Most existing map matching algorithms take a single GPS trajectory as input. We refer them as single-track map matching algorithms (SMM). They are typically formulated with the objective of minimizing the distance between the projected path on the map and the input trajectory, and of achieving some other regularization objectives, such as minimizing the length of the path. These algorithms work well when the input trajectory is densely sampled and the sampling error is small. However, their performance drops significantly when the input trajectory becomes noisy and sparse. In this case, the estimated path does not necessarily need to be close to the input trajectory, and it may not always follow or approximate the shortest path on the map.

In our work we address these issues using multi-track map matching, i.e. simultaneously matching a collection of trajectories to a map. The advantage of this approach comes again from the observation that human trajectories show a ``high degree of temporal and spatial regularity”. In the context of map matching, we have observed large amount of repeated regular structures in vehicle trajectories despite being driven by different drivers. Hence the aim of multi-track map matching is to recover the regularity patterns (i.e., frequently used road segments) among the input trajectories, and to preserve the regularity in the

matched paths. From a data-driven perspective, multi-track map matching offers additional regularization constraints that improve the map matching results of individual trajectories --- effectively using the ``wisdom of the collection'' to compensate for noise and gaps in individual trajectories. Specifically, for a set of partially overlapping trajectories, we enforce that the projected paths of their overlapping regions coincide. Such a formulation implicitly increases the sampling density of trajectories. Moreover, the overlapping parts of these trajectories are jointly determined, which improves the robustness of the map matching procedure. The multi-track idea was used in earlier work, with the assumption that all trajectories are sampled from the same underlying path. Our algorithm, designed to apply multi-track map matching on heterogeneous data, offers more practical use on large-scale map matching applications and applies (as well a benefits) from path diversity in the collection.

II. Locating Lucrative Passengers for Taxicab Drivers

In a big city like Beijing, there are more than 10,000 taxis operating every day, and the majority of taxi passengers find their taxis by standing beside the street and waiting for a vacant one to come by. Hence, for taxicab drivers, every time after they drop their previous passengers, they have to make a decision about where to search for the next passenger. When the objective of a driver is to maximize the daily income -- a natural strategy is to search within an “attractive” area where the chance of finding a passenger is high. However, just finding any passenger is not sufficient: taxi drivers prefer long trips, since they are more profitable. On the other hand, long trips may force the driver into a remote area of the city where finding the following passenger will be difficult. As every taxi driver faces this problem several times per day, he/she develops a --- perhaps unconscious --- strategy based on personal experience.

In this work we posed the question of whether a good strategy for finding a passenger can be computed from GPS trace data; such a strategy could be used, for instance, as a basic guideline for an inexperienced taxi driver. For that purpose, we model the problem of finding a lucrative passenger as a Markov Decision Process (MDP). All parameters of the MDP are obtained by analyzing a collection of GPS data of 1000 taxis over one month in Beijing, China. We compute an optimal policy for the MDP using dynamic programming; that policy can be represented as a directed acyclic graph (DAG) where an edge from location A to location B represents a recommendation to drive from A to B when looking for a passenger. In particular, a sink in the DAG is a locally optimal area to find a pickup and the taxi should stay at this location. We compute such policies for weekday daylight hours, weekday night hours, and weekends, demonstrating that the policies are changing in a meaningful way. We validate the computed policies using the same GPS data: we identify instances where the search path of the taxi drivers agrees with our proposed policy and show that such instances generate more income than the average trip.

As this example shows, when we analyze and understand trajectory collections, we end up with more than just a tool for compressing or parametrizing trajectories. Trajectories, besides being of interest in themselves, are also a tool for understanding the environment in which the motions took place and the mobile entities that traversed or generated them. A large collection of taxicab GPS traces from Beijing tells us a lot about the road structure of the city, its population hubs and how people move between them, and even about the capabilities of the driven vehicles and proclivities of the drivers themselves.

Technology Transfer